

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number  
WO 01/53312 A1(51) International Patent Classification<sup>7</sup>: C07H 21/04,  
C12N 15/11, 15/63, 15/70, 15/82, 15/85, C07K 14/00

(21) International Application Number: PCT/US00/34263

(22) International Filing Date:  
26 December 2000 (26.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/488,725	21 January 2000 (21.01.2000)	US
09/552,317	25 April 2000 (25.04.2000)	US
09/598,042	9 July 2000 (09.07.2000)	US
09/620,312	19 July 2000 (19.07.2000)	US
09/653,450	3 August 2000 (03.08.2000)	US
09/662,191	14 September 2000 (14.09.2000)	US
09/693,036	19 October 2000 (19.10.2000)	US
09/727,344	29 November 2000 (29.11.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	09/488,725 (CIP)
Filed on	21 January 2000 (21.01.2000)
US	09/552,317 (CIP)
Filed on	25 April 2000 (25.04.2000)
US	09/598,042 (CIP)
Filed on	9 July 2000 (09.07.2000)
US	09/620,312 (CIP)
Filed on	19 July 2000 (19.07.2000)
US	09/653,450 (CIP)
Filed on	3 August 2000 (03.08.2000)
US	09/662,191 (CIP)
Filed on	14 September 2000 (14.09.2000)
US	09/693,036 (CIP)
Filed on	19 October 2000 (19.10.2000)
US	09/727,344 (CIP)
Filed on	29 November 2000 (29.11.2000)

(71) Applicant (for all designated States except US): HYSEQ, INC. [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). LIU, Chenghua [CN/US]; 1125 Ranchero Way #14,

San Jose, CA 95117 (US). ASUNDI, Vinod [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). CHEN, Rui-hong [US/US]; 1031 Flying Fish Street, Foster City, CA 94404 (US). MA, Yunqing [CN/US]; 280 W. California Avenue #206, Sunnyvale, CA 94086 (US). QIAN, Xiaohong, B. [CN/US]; 3662 Tumble Way, San Jose, CA 95132 (US). REN, Feiyan [US/US]; 7703 Oak Meadow Court, Cupertino, CA 95014 (US). WANG, Dunrui [CN/US]; 932 La Palma, Milpitas, CA 95035 (US). WANG, Jian-Rui [CN/US]; 744 Stendhal Lane, Cupertino, CA 95014 (US). WANG, Zhiwei [CN/US]; 836 Alturas Avenue, B36, Sunnyvale, CA 94085 (US). WEHRMAN, Tom [US/US]; 3210 CCSR Mol Pharm, 269 W. Campus Drive, Stanford, CA 94305 (US). XU, Chongjun [CN/US]; 4918 Manitoba Drive, San Jose, CA 95130 (US). XUE, Aidong, J. [CN/US]; 1621 S. Mary Avenue, Sunnyvale, CA 94087 (US). YANG, Yonghong [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). ZHANG, Jie [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). ZHAO, Qing, A. [CN/US]; 1556 Kooser Road, San Jose, CA 95118 (US). ZHOU, Ping [CN/US]; 1461 Japaul Lane, San Jose, CA 95132 (US). GOODRICH, Ryle [US/US]; 4896 Sandy Lane, San Jose, CA 95124 (US). DRMANAC, Radoje, T. [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky, and Popco, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

## Published:

with international search report

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

[Continued on next page]

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.



WO 01/53312 A1



---

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1-1786 and 3573-5358, a mature protein coding portion of SEQ ID NO:1-1786 and 3573-5358, an active domain of SEQ ID NO:1-1786 and 3573-5358, and complementary sequences thereof.
2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.
3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.
4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.
5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.
6. A vector comprising the polynucleotide of claim 1.
7. An expression vector comprising the polynucleotide of claim 1.
8. A host cell genetically engineered to comprise the polynucleotide of claim 1.
9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.
10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

- (a) a polypeptide encoded by any one of the polynucleotides of claim 1; and
  - (b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO:1-1786 and 3573-5358.
11. A composition comprising the polypeptide of claim 10 and a carrier.
12. An antibody directed against the polypeptide of claim 10.
13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:
- a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and
  - b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.
14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:
- a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;
  - b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and
  - c) detecting said product and thereby the polynucleotide of claim 1 in the sample.
15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.
16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO:1-1786 and 3573-5358, a mature protein coding portion of SEQ ID NO:1-1786 and 3573-5358, an active domain of SEQ ID NO:1-1786 and 3573-5358, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO:1-1786 and 3573-5358, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides SEQ ID NO:1787 -3572 and 5359-7144, the mature protein portion thereof, or the active domain thereof.
21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.
22. A collection of polynucleotides, wherein the collection comprising the sequence information of at least one of SEQ ID NO:1-1786 and 3573-5358.
23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.
24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.
25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.
26. The collection of claim 22, wherein the collection is provided in a computer-readable format.
27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.
28. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

TABLE 2

SEQ ID NO:	ACCESSION NUMBER	SPECIES	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
			VDJ region		
344	U10281	Sus scrofa	gastric mucin	279	24
345	AK000404	Homo sapiens	unnamed protein product	1177	99
346	L22557	Rattus norvegicus	calmodulin-binding protein	1949	84
347	L22557	Rattus norvegicus	calmodulin-binding protein	2363	91
348	AL049481	Arabidopsis thaliana	AIG1-like protein	316	30
350	AJ251516	Mus musculus	cysteine and histidine-rich protein	1460	99
351	AK024477	Homo sapiens	FLJ00070 protein	1773	100
352	U50133	Homo sapiens	ankyrin	502	33
353	AK000625	Homo sapiens	unnamed protein product	721	100
354	AF161420	Homo sapiens	HSPC302	2623	97
355	AJ010014	Homo sapiens	M96A protein	1269	47
356	AF151029	Homo sapiens	HSPC195	941	91
357	AL022327	Homo sapiens	dJ355C18.1 (KIAA0027)	1911	100
358	W78128	Homo sapiens	Human secreted protein encoded by gene 3 clone HOSBI96.	1117	100
359	X03414	Drosophila melanogaster	Kr polypeptide	316	45
360	AF151079	Homo sapiens	HSPC245	643	100
361	Y53886	Homo sapiens	A suppressor of cytokine signalling protein designated HSCOP-6.	530	41
362	AF254741	Drosophila melanogaster	Centaurin Gamma 1A	681	46
363	AF213465	Homo sapiens	dual oxidase	2016	100
364	AF181562	Homo sapiens	proSAAS	1319	100
365	AF181562	Homo sapiens	proSAAS	1024	99
366	U73200	Mus musculus	pl16Rip	884	82
367	AF263744	Homo sapiens	erbB2-interacting protein ERBIN	4973	99
368	U37501	Mus musculus	laminin alpha 5 chain	5867	72
369	AF043695	Caenorhabditis elegans	similar to the protein phosphates 2c family	549	36
370	Y73440	Homo sapiens	Human secreted protein clone yj23_1 protein sequence SEQ ID NO:102.	1484	99
371	AF272833	Homo sapiens	misato	2869	97
372	AF198454	Homo sapiens	epithelial protein lost in neoplasm beta	3927	100
373	Y73345	Homo sapiens	HTRM clone 438283 protein sequence.	273	80
374	AF169017	Homo sapiens	formiminotransferase cyclodeaminase	2717	98
375	A95106	unidentified	RED ALPHA	1202	99
376	W74828	Homo sapiens	Human secreted protein encoded by gene 100 clone HLQAB52.	1012	99
377	Y32131	Homo sapiens	Human LYST-2 protein.	3556	99
378	M14912	Homo sapiens	pol	132	86
379	AF090934	Homo sapiens	PRO0518	382	100
380	X66363	Homo sapiens	serine/threonine protein kinase	2499	100
381	Y41699	Homo sapiens	Human PRO703 protein sequence.	2362	100
382	AF174498	Homo sapiens	GR AF-1 specific protein phosphatase	7008	98
383	U64608	Caenorhabditis elegans	coded for by C. elegans cDNA yk173c12.5	246	36
384	U50133	Homo sapiens	ankyrin	502	33
385	AJ238520	Homo sapiens	putative transcription factor-like nuclear regulator	4123	97





SEQ ID NO:	PFAM NAME	DESCRIPTION	p-value	PFAM SCORE
335	vwa	von Willebrand factor type A domain	2.3e-07	37.9
339	ras	Ras family	7.8e-07	-59.1
340	zf-C2H2	Zinc finger, C2H2 type	8.2e-64	225.4
342	zf-C2H2	Zinc finger, C2H2 type	2.4e-85	297.0
343	ig	Immunoglobulin domain	0.0005	18.0
346	pkinase	Eukaryotic protein kinase domain	6.5e-65	229.1
347	pkinase	Eukaryotic protein kinase domain	6.5e-65	229.1
351	EGF	EGF-like domain	8.5e-20	79.2
352	ank	Ank repeat	2.5e-101	350.0
354	TBC	TBC domain	5.1e-15	63.3
355	PHD	PHD-finger	3.2e-07	37.4
358	DUF6	Integral membrane protein DUF6	0.033	15.8
359	zf-C2H2	Zinc finger, C2H2 type	7.4e-20	79.4
361	ank	Ank repeat	6.6e-34	126.1
362	ArfGap	Putative GTP-ase activating protein for Arf	4.7e-53	189.7
363	efhand	EF hand	5.4e-10	46.6
367	LRR	Leucine Rich Repeat	8.8e-44	158.9
368	laminin_G	Laminin G domain	1.5e-33	121.7
369	PP2C	Protein phosphatase 2C	5.3e-20	73.9
372	LIM	LIM domain containing proteins	9.9e-15	57.1
373	KRAB	KRAB box	4.8e-23	90.0
376	ion_trans	Ion transport protein	2.9e-09	-4.2
377	Beach	Beige/BEACH domain	4.9e-208	704.5
380	pkinase	Eukaryotic protein kinase domain	1.6e-94	327.5
381	AMP-binding	AMP-binding enzyme	1.4e-07	-140.3
382	HECT	HECT-domain (ubiquitin-transferase).	1.3e-07	-13.5
384	ank	Ank repeat	2.5e-101	350.0
386	ig	Immunoglobulin domain	9.5e-06	23.6
388	zf-C2H2	Zinc finger, C2H2 type	1.7e-42	154.6
389	ig	Immunoglobulin domain	2.8e-15	54.3
390	mito_carr	Mitochondrial carrier proteins	3.5e-67	233.2
392	TPR	TPR Domain	6.1e-17	69.7
393	SH3	SH3 domain	3.5e-09	43.9
394	AAA	ATPases associated with various cellular act	4.1e-21	83.6
396	spectrin	Spectrin repeat	2.1e-67	237.3
397	zf-C2H2	Zinc finger, C2H2 type	0.0066	23.1
399	fn3	Fibronectin type III domain	4.1e-102	352.6
400	WD40	WD domain, G-beta repeat	0.00049	26.8
401	E1_dehydrog	Dehydrogenase E1 component	3e-119	409.6
402	fn3	Fibronectin type III domain	0	1719.6
404	LRR	Leucine Rich Repeat	2.1e-10	48.0
405	cadherin	Cadherin domain	8.1e-81	281.9
406	zf-CXXC	CXXC zinc finger	5e-15	63.4
410	RhoGEF	RhoGEF domain	1.1e-23	92.1
411	F-box	F-box domain.	4.2e-06	33.7
412	SNF2_N	SNF2 and others N-terminal domain	5.8e-16	61.6
415	CPSase_L_cha in	Carbamoyl-phosphate synthase (CPSase)	1.5e-172	586.6
418	LRR	Leucine Rich Repeat	3.8e-24	93.6
419	DENN	DENN (AEX-3) domain	2e-58	207.5
420	RasGEF	RasGEF domain	8.1e-43	155.7
421	ank	Ank repeat	1.4e-153	523.7
424	G-patch	G-patch domain	1e-19	78.9
425	pkinase	Eukaryotic protein kinase domain	2.2e-31	117.1
426	Plexin_repea t	Plexin repeat	0.0023	24.6
427	Plexin_repea	Plexin repeat	0.0023	24.6



SEQ ID NO: of full- length nucleotide sequence	SEQ ID NO: of full- length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Priority docket number_ corresponding SEQ ID NO: in priority application	SEQ ID NO: in U.S.S.N. 09/488,725
308	2094	3880	5666	784CIP2_312	9494
309	2095	3881	5667	784CIP2_313	9512
310	2096	3882	5668	784CIP2_314	9632
311	2097	3883	5669	784CIP2_315	9661
312	2098	3884	5670	784CIP2_316	9664
313	2099	3885	5671	784CIP2_317	9691
314	2100	3886	5672	784CIP2_318	9700
315	2101	3887	5673	784CIP2_319	9716
316	2102	3888	5674	784CIP2_320	9721
317	2103	3889	5675	784CIP2_321	9870
318	2104	3890	5676	784CIP2_322	9887
319	2105	3891	5677	784CIP2_323	9923
320	2106	3892	5678	784CIP2_324	9938
321	2107	3893	5679	784CIP2_325	9964
322	2108	3894	5680	784CIP2_326	10007
323	2109	3895	5681	784CIP2_327	10009
324	2110	3896	5682	784CIP2_328	10046
325	2111	3897	5683	784CIP2_329	10156
326	2112	3898	5684	784CIP2_330	10276
327	2113	3899	5685	784CIP2_331	10283
328	2114	3900	5686	784CIP2B_1	152
329	2115	3901	5687	784CIP2B_2	167
330	2116	3902	5688	784CIP2B_3	205
331	2117	3903	5689	784CIP2B_4	210
332	2118	3904	5690	784CIP2B_5	225
333	2119	3905	5691	784CIP2B_6	226
334	2120	3906	5692	784CIP2B_7	264
335	2121	3907	5693	784CIP2B_8	268
336	2122	3908	5694	784CIP2B_9	293
337	2123	3909	5695	784CIP2B_10	293
338	2124	3910	5696	784CIP2B_11	293
339	2125	3911	5697	784CIP2B_12	302
340	2126	3912	5698	784CIP2B_13	311
341	2127	3913	5699	784CIP2B_14	352
342	2128	3914	5700	784CIP2B_15	358
343	2129	3915	5701	784CIP2B_16	368
344	2130	3916	5702	784CIP2B_17	393
345	2131	3917	5703	784CIP2B_18	477
346	2132	3918	5704	784CIP2B_19	508
347	2133	3919	5705	784CIP2B_20	508
348	2134	3920	5706	784CIP2B_21	515
349	2135	3921	5707	784CIP2B_22	578
350	2136	3922	5708	784CIP2B_23	588
351	2137	3923	5709	784CIP2B_24	591
352	2138	3924	5710	784CIP2B_25	593
353	2139	3925	5711	784CIP2B_26	594
354	2140	3926	5712	784CIP2B_27	619
355	2141	3927	5713	784CIP2B_28	620
356	2142	3928	5714	784CIP2B_29	654
357	2143	3929	5715	784CIP2B_30	692
358	2144	3930	5716	784CIP2B_31	753
359	2145	3931	5717	784CIP2B_32	758
360	2146	3932	5718	784CIP2B_33	787
361	2147	3933	5719	784CIP2B_34	833
362	2148	3934	5720	784CIP2B_35	838
363	2149	3935	5721	784CIP2B_36	870
364	2150	3936	5722	784CIP2B_37	891
365	2151	3937	5723	784CIP2B_38	891
366	2152	3938	5724	784CIP2B_39	921
367	2153	3939	5725	784CIP2B_40	924
368	2154	3940	5726	784CIP2B_41	932
369	2155	3941	5727	784CIP2B_42	942

